FIRST EXPERIMENTS WITH RUE AUTOMATED DEDUCTION

Vincent J. Digricoli

The Courant Institute and Hofstra University
251 Mercer Street, New York, N.Y. 10012

ABSTRACT

RUE resolution represents a reformulation of binary resolution so that the basic rules of inference (RUE and NRF) incorporate the axioms of equality. An RUE theorem prover has been implemented and experimental results indicate that this method represents a significant advance in the handling of equality in resolution.

A. Introduction

In (1) the author presented the complete theory of Resolution by Unification and Equality which incorporates the axioms of equality into two inference rules which are sound and complete to prove E-unsatisfiability. Our purpose here is to present systematically the results of experiments with an RUE theorem prover.

The experiments chosen were those of McCharen, Overbeek and Wos (2), and in particular we are interested in comparing the results achieved by these two theorem provers.

In MOW, the equality axioms were used explicitly for all theorems involving equality and apparently no use was made of paramodulation. In RUE, where proofs are much shorter, the inference rules themselves make implicit use of the equality axioms which do not appear in a refutation and also no use of paramodulation is made. Both systems are pure resolution-based systems.

Before considering the experiments, we first review and summarize the theory of resolution by unification and equality as presented in (1). There we define the concept of a disagreement set, the inference rules RUE and NRF, the notion of viability, the RUE unifying substitution and an equality restriction which inhibits redundant inferences. Here we simply introduce the concept of a disagreement set and define the rules of inference.

A disagreement set of a pair of terms (t_1,t_2) is defined in the following manner :

"If (t_1,t_2) are identical, the empty set is the only disagreement set and if (t_1,t_2) differ, the set of one pair $\{(t_1,t_2)\}$ is the <u>origin disagreement set</u>. Furthermore, if t_1 has the form $f(a_1,\ldots,a_k)$ and t_2 the form $f(b_1,\ldots,b_k)$, then the set of pairs of corresponding arguments which are not identical is the <u>topmost disagreement set</u>. Finally, if D is a disagreement set of (t_1,t_2), then D' obtained from D by replacing any member of D by the elements of one of its disagreement sets, is also a disagreement set of (t_1,t_2)."

In the simple example :

$$t_1 = f(a, g(b,h(c)))$$
$$t_2 = f(a', g(b',h(c')))$$

besides the origin disagreement, there are the disagreement sets :

$$D_1 = \{(a,a'),\ (g(b,h(c)),g(b',h(c')))\}$$
$$D_2 = \{(a,a'),\ (b,b'),\ (h(c),h(c'))\}$$
$$D_3 = \{(a,a'),\ (b,b'),\ (c,c')\}$$

This definition merely defines all possible ways of proving $t_1=t_2$, i.e. we can prove $t_1=t_2$ by proving equality in every pair of any one disagreement set. An input clause set, for example, may imply equality in D_1 but not in D_2 or D_3. Or it may most directly prove $t_1=t_2$ by proving equality in D_3.

We proceed to define a <u>disagreement set of complementary literals</u> :

$$P(s_1,\ldots,s_n) \qquad \bar{P}(t_1,\ldots,t_n)$$

as the union of disagreement sets :

$$D = \bigcup_{i=1,n} D_i$$

where D_i is a disagreement set of (s_i,t_i).

We see immediately that :

$$P(s_1,\ldots,s_n) \wedge \bar{P}(t_1,\ldots,t_n) \longrightarrow D$$

where D now represents the disjunction of inequalities specified by a disagreement set of $P,\bar{P}$, and furthermore, that :

$$f(a_1,\ldots,a_k) \neq f(b_1,\ldots,b_k) \longrightarrow D$$

where D is the disjunction of inequalities specified by a disagreement set of $f(a_1,\ldots,a_k)$, $f(b_1,\ldots,b_k)$. For example,

$$P(f(a,g(b,h(c)))) \wedge \bar{P}(f(a',g(b',h(c'))))$$
$$\longrightarrow a\neq a' \wedge b\neq b' \wedge c\neq c' .$$

The reader is invited to read (1) which states the complete theory of RUE resolution with many examples. Our primary concern here is to discuss experiments with an RUE theorem prover and to begin to assess the effectiveness of this inference system.

B. Experiments

Our experiments deal with Boolean Algebra and we are asked to prove from the eight axioms :

A1 : $x + 0 = x$
A2 : $x * 1 = x$
A3 : $x + \bar{x} = 1$
A4 : $x * \bar{x} = 0$
A5 : $x(y+z) = xy + xz$
A6 : $x + yz = (x+y)(x+z)$
A7 : $x + y = y + x$
A8 : $x * y = y * x$

(we are denoting logical or by +, logical and by * or juxtaposition, and negation by overbar),

the following theorems :

T1 : $\bar{0} = 1$

T2 : $x + 1 = 1$

T3 : $x * 0 = 0$

T4 : $x + xy = x$

T5 : $x(x+y) = x$

T6 : $x + x = x$

T7 : $x * x = x$

T8 : $(x+y)+z = x+(y+z)$

T9 : $(x*y)*z = x*(y*z)$

T10 : the complement of x is unique
$(x*a=0)(x+a=1)(x*b=0)(x+b=1) \longrightarrow a = b$

T11 : $\bar{\bar{x}} = x$

T12 : $\overline{x + y} = \bar{x} * \bar{y}$ De Morgan's Law I

T13 : $\overline{x * y} = \bar{x} + \bar{y}$ De Morgan's Law II

These theorems are stated in the order of increasing complexity of proof, with $\bar{0}= 1$ being trivially easy for a human to prove and De Morgan's Laws being very difficult for a human to deduce from the axioms.

George and Garrett Birkhoff have a paper on the above proofs published in the Transactions of the American Mathematical Society (3) and Halmos comments on the significantly difficult character of the proofs in his Lectures on Boolean Algebras (4).

The following is a machine deduced, five step RUE refutation which proves $x*0 = 0$:

$a*0 \neq 0$
$\vdash$ $x*\bar{x} = 0$ $\sigma = \{a/x\}$
$a*\bar{a} \neq a*0$
$\vdash$ $x(y+z) = xy + xz$ $\sigma = \{a/x\}$
$y+z \neq \bar{a}$ $ay+az \neq a*0$
$\vdash$ $x+0=x$ $\sigma = \{\bar{a}/y, 0/z, \bar{a}/x\}$
$a*\bar{a} + a*0 \neq a*0$
$\vdash$ $0 + x = x$ $\sigma = \{a*0/x\}$
$0 \neq a*\bar{a}$
$\vdash$ $0 = x*\bar{x}$ $\sigma = \{a/x\}$
$\square$

The above experiments together with many others (dealing with group theory, ring theory, geometry, Henken Models, set theory and program verification) were proposed as benchmarks by McCharen,Overbeek and Wos, who in (2) published the results of their own experiments.

We here tabulate the comparative performance of the RUE and MOW theorem provers on the above theorems. The MOW theorem prover uses binary resolution with explicit use of the equality axioms and is implemented in Assembly language on the IBM System 370-Model 195. Great effort was made to enhance the efficiency of their theorem prover and this is described in (2). The RUE theorem prover, on the other hand, represents a first implementation in PL1 on a CDC 6600 machine which is much slower than the Model 195.

In the experiments each theorem is treated as an independent problem and cannot use earlier theorems as lemmas, so that for example in proving associativity (T8), we need to prove (T2,T3,T4,T5) as sub-theorems. The total number of unifications performed is suggested as the primary measure of comparison rather than time. The comparative results are given in Table 1.

From T1 to T7, The RUE theorem prover was very successfull, but at T8 (associativity) results have yet to be obtained since refinements in the heuristic pruning procedure are required and are being developed with the expectation that more advanced results will be available at the conference.

RUE represents one of several important methods for handling equality in resolution and it is important to emphasize that it is a complete method whose power is currently being tested in stand-alone fashion. However, it is not precluded that we can combine this method with other techniques such as demodulation,paramodulation and reduction theory to achieve a mutually enhanced effect.

TABLE 1.

THEOREM	TOTAL NUMBER OF UNIFICATIONS RUE	: MOW	TIME (SECONDS) RUE	: MOW	LENGTH OF PROOF RUE	: MOW
T1 $\bar{0} = 1$	77	26,702	2.9	16.2	2	*
T2 $x + 1 = 1$	688	46,137	10.3	28.5	7	
T3 $x * 0 = 0$	676	46,371	10.1	27.5	7	
T4 $x + xy = x$	3,152	see below	51.4	see below	12	
T5 $x(x+y) = x$	3,113	" "	51.5	" "	12	
T4∧T5	6,265[1]	286,902	102.9	57.0	24	
T6 $x + x = x$	2,181	see below	41.6	see below	13	
T7 $x * x = x$	2,145	" "	41.4	" "	13	
T6∧T7	4,326[1]	105,839	83.0	60.6	26	
T8 $(x+y)+z = x+(y+z)$	IP	413,455	IP	177.2		
T9 $(x*y)*z = x*(y*z)$	IP	NPR				
T10 $(x*a=0)\ (x+a=1)$ $(x*b=0)\ (x+b=1) \rightarrow a=b$	IP	NPR				
T11 $\bar{\bar{x}} = x$	IP	NPR				
T12 $\overline{x + y} = \bar{x} * \bar{y}$	IP	NPR				
T13 $\overline{x * y} = \bar{x} + \bar{y}$	IP	NPR				

IP : IN PROCESS

NPR : NO PUBLISHED RESULT

Note 1 : To prove the double theorem, T4∧T5, $x+xy=x \wedge x(x+y)=x$, we add the negated theorem as a single clause, $a+ab\neq a \vee a(a+b)\neq a$, to the input clause set. It is evident that the erasure of these two literals in a refutation decomposes into two independent subproblems since no variables appear in the clause. Hence, the refutations for $a+ab\neq a$ and $a(a+b)\neq a$ obtained in separate experiments T4,T5 can be concatenated and the results of these experiments simply summed which is what we have done to state the RUE results for the double theorem. The same holds true for T6∧T7.

* The estimated length of MOW proofs with the equality axioms is twice as long as corresponding RUE proofs.

The completion of these Boolean experiments together with other experiments with a more fully delineated comparative analysis with systems other than MOW represents work the author will report upon in the future.

REFERENCES :

1. "Automatic Deduction and Equality" by Vincent J. Digricoli, Proceedings of the Oct., 1979, Annual Conference of the ACM, 240-250.

2. "Problems and Experiments for and with Automated Theorem-Proving Programs", by McCharen, Overbeek and Wos, IEEE Transactions on Computers, Vol C-25,No.8,August 1976.

3. "Distributive Postulates for Systems Like Boolean Algebras", Transactions of the American Mathematical Society, Volume 60, July-Dec. 1946.

4. "Lectures on Boolean Algebras", by Paul R. Halmos, Section 2, pages 3-5, D.Van Nostrand.